AF /

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Appellants:	Franz et al.	Examiner: J. Jackson
Serial No:	09/782,434	Group Art Unit: 2655
Filed:	February 13, 2001	Docket: YOR9-2001-0011US1 (8728-481)
For:	DYNAMIC LANGUAGE MODEL MIXTURES WITH HISTORY-BASED BUCKETS	

APPEAL BRIEF

This is an Appeal from the Final Office Action mailed June 9, 2004 (Paper No. 4), finally rejecting claims 1 and 3-24. Applicants appeal pursuant to the Notice of Appeal filed on November 3, 2004 and submit this appeal brief.

Appeal from Group 2655

F. Chau & Associates, LLC
130 Woodbury Road
Woodbury, NY 11797
TEL: (516) 692-8888
FAX: (516) 692-8889
Attorneys for Appellants

TABLE OF CONTENTS

1. **Real Party in Interest**

The real party in interest is INTERNATIONAL BUSINESS MACHINES CORPORATION, the assignee of the entire right, title and interest in and to the subject application by virtue of an assignment of record.

2. **Related Appeals and Interferences**

None.

3. **Status of Claims**

Claims 1 and 3-24 are pending, stand rejected and are under appeal.

A copy of the claims 1 and 3-24 as pending is presented in the Claims Appendix.

4. **Status of Amendments**

An amendment to claims 3, 4 and 7 were proposed after Final Rejection (Paper No. 4) solely to overcome an objection (*i.e.*, claims 3, 4 and 7 depended from cancelled claim 2). Although the corresponding Advisory Action (Paper No. 20040929) did not enter the amendment, all objections were removed, as stated in the Advisory Action. It is believed that the decision not to enter the amendment was in error; thus, the Claims Appendix lists claims 3, 4, 7 correctly depending from claim 1, instead of claim 2.

The remaining claims were not amended after Final Rejection.

5. Summary of Claimed Subject Matter

In independent claim 1, a method for combining language model scores generated by at least two language models, in an automatic speech recognition system ("ASR"), is presented. The method includes the following steps. A list of most likely words for a current word in a word sequence uttered by a speaker and acoustic scores corresponding to the most likely words are generated. (Figure 2, # 210, #212; Figure 3, #310). Language model scores for each of the most likely words in the list, for each of the at least two language models, are computed. (Figure 2, #214a-#214n; Figure 3, #312). A set of coefficients to be used to combine the language model scores of each of the most likely words in the list, based on a context of the current word, are respectively and dynamically determinined. (Figure 2, #215; Figure 3, #314). The set of coefficients are respectively and dynamically determined by dividing text data for training a plurality of sets of coefficients into partitions, depending on word counts corresponding to each of the at least two language models, and for each of the most likely words in the list, by dynamically selecting the set of coefficients from among the plurality of sets of coefficients so as to maximize the likelihood of the text data with respect to the at least two language models. (Figure 4). The language model scores of each of the most likely words in the list are respectively combined to obtain a composite score for each of the most likely words in the list, using the set of coefficients determined therefor. (Figure 3, #316).

In independent claim 11, a method for combining language model scores generated by at least two language models comprised in an Automatic Speech Recognition ("ASR") system is provided. The method includes the following steps. A list of most likely words for a current word in a word sequence uttered by a speaker and acoustic scores

corresponding to the most likely words are generated. (Figure 2, # 210, #212; Figure 3, #310). Language model scores for each of the most likely words in the list, for each of the at least two language models, are computed. (Figure 2, #214a-#214n; Figure 3, #312). A weight vector to be used to combine the language model scores of each of the most likely words in the list based on a context of the current word is respectively and dynamically determined. (P. 16, lines 5-10). The weight vector includes n-weights, wherein n equals a number of language models in the system, and each of the n-weights depends upon history n-gram counts. (P. 20, line 10-18). The language model scores of each of the most likely words in the list are respectively combined to obtain a composite score for each of the most likely words in the list, using the weight vector determined therefor. (Figure 3, #316).

In independent claim 19, a combining system for combining language model scores generated by at least two language models comprised in an Automatic Speech Recognition ("ASR") system is provided The ASR system includes a fast match for generating a list of most likely words for a current word in a word sequence uttered by a speaker and acoustic scores corresponding to the most likely words. (Figure 2, #212). The combining system includes the following:

(1) a language model score computation device adapted to compute language model scores for each of the most likely words in the list, for each of the at least two language models; (Figure 2, #214a-#214n)

(2) a selection device adapted to respectively and dynamically select a weight vector to be used to combine the language model scores of each of the most likely words in the list based on a context of the current word, the weight vector comprising n-weights,

wherein n equals a number of language models in the system, and each of the n-weights depends upon history n-gram counts; and (Figure 2, #215)

(3) a combination device adapted to respectively combine the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the weight vector selected therefor. (Figure 2, #216).

6. Grounds of Rejection to be Reviewed on Appeal

A. Claims 1, 3, 5-13 and 15-24 stand rejected under 35 U.S.C. § 102(e) as being anticipated by Gillick et al. (U.S. Patent No. 6,167,377) (hereinafter "Gillick").

7. Argument

A. Introduction

A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. *See Glaxo Inc. v. Novopharm Ltd.*, 52 F.3d 1043, 1047, 34 USPQ2d 1565, 1567 (Fed. Cir. 1995). In other words, there must be no difference between the claimed invention and the reference disclosure, as viewed by a person of ordinary skill in the field of the invention. *See Scripps Clinic & Research Found. v. Genentech Inc.*, 927 F.2d 1565, 1576, 18 USPQ2d 1001, 1010 (Fed. Cir. 1991). An anticipation rejection cannot be predicated on an ambiguous reference. Rather, statements and drawings in a reference relied on to prove anticipation must be so clear and explicit that those skilled in the art will have no difficulty

in ascertaining their meaning. *See In re Turlay*, 304 F.2d 893, 899, 134 USPQ 355, 360 (CCPA 1962).

It is respectfully submitted that the Examiner has failed to show that the reference Gillick describes each and every limitation in the rejected claims. In particular, the reference Gillick fails to describe "determining a set of coefficients to be used to combine the language model scores" and "dividing text data for training a plurality of sets of coefficients," as claimed in claim 1. Further, the reference Gillick fails to describe "the weight vector comprising n-weights, wherein n equals a number of language models in the system, and each of the n-weights depends upon history n-gram counts," as claimed in claims 11 and 19. For the reasons set forth below, Appellants respectfully request that the claim rejections under 35 U.S.C. § 102(e) be reversed.

B. Claims 1, 3, 5-13 and 15-24 stand rejected under 35 U.S.C. § 102(e) as being anticipated by Gillick et al. (U.S. Patent No. 6,167,377).

(i). Gillick fails to describe "determining a set of coefficients to be used to combine the language model scores, based on a context of the current word," as claimed in claim 1.

The Examiner cites col. 17, lines 39-41 of Gillick as anticipating "determining a *set of coefficients* to be used to combine the language model scores, *based on a context of the current word*," as claimed in claim 1. (Paper no. 4, p. 4). The recited portion of Gillick simply states that "[t]he first language model is a bigram model that indicates the frequency with which a word occurs in the context of a preceding word." What the first language model is is entirely unrelated to *combining* the language model scores. The recited portion of Gillick clearly does not disclose "determining *a set of coefficients*,"

much less "determining a set of coefficients to *combine* the language model scores *based on a context of the current word*," also as claimed in claim 1.

The Advisory Action further cites "frames of a parameter" and "figure 2, element 210" of Gillick as disclosing "determining a set of coefficients." The recited portion of Gillick describes "frames of parameters 210 that represent the frequency content of an utterance." (Gillick, col. 3, lines 34-36). Gillick states that "[i]n a frame-based system, a processor divides a signal descriptive of the speech to be recognized into a series of digital frames, each of which corresponds to a *small time increment of the speech*." Clearly, the recited portion of Gillick does not disclose "determining a set of coefficients." Even assuming, *arguendo*, that Gillick discloses "determining a set of coefficents," the "frames of a parameters," as described in Gillick, are *not* used to combine language model scores: "The recognizer 215 then combines the scores produced by the language models *using interpolation weights* to produce a combined language model score for each word (step 1310)."

It should be noted that the Examiner is inconsistent in many rejections; in fact, the Examiner's rejections, in many cases, simply do not make logical sense. For example, as described above, the Examiner attempts to cite figure 2, element 210 of Gillick, which describes *frames*, to anticipate "determining a set of coefficients." The Examiner also attempts to cite col. 15, lines 7-13 of Gillick as anticipating "training a plurality of sets of coefficients," also as claimed in claim 1. (Paper no. 4, p. 4). Now assuming, *arguendo*, that the Examiner correctly anticipates "determining a set of coefficients," it *should* follow that a citation to "training a plurality of sets of coefficients" should train *the frames* (*i.e.*, what the Examiner considers to describe "a set of coefficients"). However, the recited

portion of Gillick states that "[t]he enrollment program collects acoustic information from a user and *trains or adapts a user's models* based on that information." (col. 15, lines 9-11). A user's models is patentably distinguishable from the frames, as described in Gillick. Clearly, the Examiner's own rejections are without merit.

Because Gillick does not describe each and every limitation of claim 1, it is respectfully asserted that no *prima facie* case of anticipation has been made out. Accordingly, the rejection of claims 1 and 3-10 should be reversed.

(ii). **Gillick fails to describe "determining a weight vector to be used to combine the language model scores of each of the most likely words in the list based on a context of the current word," as claimed in claims 11 and 19.**

The Examiner cites col. 17, lines 39-41 of Gillick as anticipating "determining *a weight vector* to be used to combine the language model scores of each of the most likely words in the list *based on a context of the current word*," as claimed in claims 11 and 19. (Paper no. 4, p. 7). As described in Part (ii) above, the recited portion of Gillick describes *frames* (*i..e.*, small increments of digital speech), which are entirely unrelated to to "weight vectors," as claimed in claims 11 and 19. Further, even assuming, *arguendo*, that "frames" correctly anticpates "weight vectors," Gillick does not describe using frames "to combine language model scores," as claimed in claim 1.

Because Gillick does not describe each and every limitation of claims 11 and 19, it is respectfully asserted that no *prima facie* case of anticipation has been made out. Accordingly, the rejection of claims 11-24 should be reversed.

(iii). <u>Gillick fails to describe "dividing text data for training a plurality of sets of coefficients into partitions, depending on word counts corresponding to each of the at least two language models," as claimed in claim 1.</u>

The Examiner incorrectly argues that because "<u>Gillick</u> discloses dividing the *spoken* utterance," <u>Gillick</u> also discloses "dividing *text* data for training a plurality of sets of coefficients," as claimed in claim 1. (Paper no. 4, p. 2). The Examiner's own admission in paper no. 4 that <u>Gillick</u> does not disclose "dividing text data," as claimed in claim 1, is sufficient evidence to reverse the anticipation rejection to claim 1. Further, the Examiner does not address the Applicants original argument that a "spoken utterance" is clearly not anticipated by "*text* data," even in its broadest *reasonable* interpretation. In particular, *text* is clearly distinguishable from a *spoken* utterance. Claim 1 itself even distinguishes between text data and a spoken utterance. The concept of the "spoken utterance" is claimed in "a word sequence *uttered by a speaker*," as claimed in claim 1.

Further, the Examiner cites various, unrelated portions of <u>Gillick</u> as anticipating "dividing text data *for training a plurality of sets of coefficients*," also as claimed in claim 1. (Paper no. 4, p. 4). In particular, the disparate citations to col. 1, lines 8-13, which is a portion of the background, and col. 15, 7-13, which describes training or adapting a user's models, are not explained by the Examiner. The Examiner clearly has not established *prima facie* anticipation of the recited portion of claim 1.

Nevertheless, as was argued in part (i) above, <u>Gillick</u> does not disclose anything remotely related to "determining a set of coefficients." <u>Gillick</u> proposes an entirely different and unrelated method for combining language model scores. Thus, even assuming, *arguendo*, that dividing spoken utterances, as disclosed in <u>Gillick</u>, somehow anticipates "dividing text data," <u>Gillick</u> clearly does not anticipate "dividing text data *for*

training a plurality of sets of coefficients." Similarly, it follows that Gillick also does not anticipate "dividing text data *for training a plurality of weight vectors*," as claimed in claims 12 and 20.

Because Gillick does not describe each and every limitation of claims 1, 12 and 20, it is respectfully asserted that no *prima facie* case of anticipation has been made out. Accordingly, the rejection of claims 1, 3-10, 12, 14, 16, 20, 21 and 23 should be reversed.

(iv). **Gillick fails to describe "determining a weight vector,...the weight vector comprising n-weights, wherein n equals a number of language models in the system, and each of the n-weights depends upon history n-gram counts," as claimed in claims 11 and 19.**

The Examiner states that "[a]though [col. 16, lines 42-44 of Gillick] does not specifically disclose that each of the n-weights depend on n-gram history counts, Gillick does teach n-gram's being the number of occurrences of the given n-gram (word frequency)." (Paper no. 4, p. 2). The Examiner seemingly and inaccurately has "*history n-gram counts*," as claimed in claims 11 and 19, confused with counts of a given n-gram. One skilled in the art would *not* make such an error in view of the disclosure.

As shown in page 3, lines 1-2 of the Specification, the "count of a given n-gram is the *number of occurrences* of the given n-gram in the corpus (word frequency)." However, a "*history* n-gram" refers to the history (*i.e.*, the previous words) of the current word being determined. (Specification, p. 20, lines 10-22). Thus, in the example shown on p. 20, lines 10-22 of the Specification, a trigram model (w_1, w_2, w_3) is described throughout the recited portion of the Specification to have a bigram *history* (w_1, w_2) for determining the current word w_3. Further, given a trigram (w_1, w_2, w_3), one can compute both

the count of the given n-gram (*i.e.*, a trigram; w_1, w_2, w_3) and the count of the history n-gram (*i.e.*, the bigram; w_1, w_2)

The difference between history n-gram counts and counts of a given n-gram is significant. The presently claimed invention makes the *novel* observation that the history count, which is very different from counts of a given n-gram, is important when combining or smoothing language models. For example, consider a first language model trained on a corpus of Reuters news stories and a second language model trained on a corpus of issued patents. One may consider the first language model to be more accurate for news-like material, and the second language model to be more accurate for technical material. This means that the "weight vector," as claimed in claims 11 and 19, can be biased towards the first language model for a history sequence typically found in news material (*e.g.*, "Dow Jones"), and biased towards the second language model for a history sequence of technical materia (*e.g.*, "computer network"), considering trigram models. Gillick simply does not consider the above.

Because Gillick does not describe each and every limitation of claim 11 and 19, it is respectfully asserted that no *prima facie* case of anticipation has been made out. Accordingly, the rejection of claims 11-24 should be reversed.

E. CONCLUSION

Each and every element of the claimed invention is not described by the teachings of the applied prior art reference. The Examiner has failed to establish a *prima facie* case of anticipation of the presently claimed invention under 35 U.S.C. § 102(e) over Gillick for at least the reasons noted above. Accordingly, it is respectfully requested that the Board

reverse the rejection of claims 1 and 3-24 under 35 U.S.C. § 102(e).

Respectfully submitted,

By: ______________________
Koon Hon Wong
Reg. No. 48,459
Attorney for Appellants

F. Chau & Associates, LLC
130 Woodbury Road
Woodbury, NY 11797
TEL: (516) 692-8888
FAX: (516) 692-8889
Attorneys for Appellants

Claims Appendix

1. In an Automatic Speech Recognition (ASR) system having at least two language models, a method for combining language model scores generated by at least two language models, said method comprising the steps of:

generating a list of most likely words for a current word in a word sequence uttered by a speaker, and acoustic scores corresponding to the most likely words;

computing language model scores for each of the most likely words in the list, for each of the at least two language models;

respectively and dynamically determining a set of coefficients to be used to combine the language model scores of each of the most likely words in the list, based on a context of the current word; and

respectively combining the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the set of coefficients determined therefor;

wherein said determining step comprises the steps of:

dividing text data for training a plurality of sets of coefficients into partitions, depending on word counts corresponding to each of the at least two language models; and

for each of the most likely words in the list, dynamically selecting the set of coefficients from among the plurality of sets of coefficients so as to maximize the likelihood of the text data with respect to the at least two language models.

2. (Cancelled).

3. The method according to claim 1, wherein the at least two language models comprise a first and a second language model, and said dividing step comprises the step of grouping, in a same partition, word triplets $w_1w_2w_3$ which have a count for the word pair w_1w_2 in the first language model greater than the count for the word pair w_1w_2 in the second language model.

4. The method according to claim 1, wherein said selecting step comprises the step of applying the Baum Welch iterative algorithm to the plurality of sets of coefficients.

5. The method according to claim 1, further comprising the step of, for each of the most likely words in the list, combining an acoustic score and the composite score to identify a group of most likely words to be further processed.

6. The method according to claim 1, wherein the group of most likely words contains less words than the list of most likely words.

7. The method according to claim 1, wherein the partitions are independent from the at least two language models.

8. The method according to claim 1, further comprising the step of representing the set of coefficients by a weight vector comprising n-weights, where n equals a number of language models in the system.

9. The method according to claim 1, wherein said combining step comprises the steps of:

for each of the most likely words in the list,

multiplying a coefficient corresponding to a language model by a language model score corresponding to the language model to obtain a product for each of the at least two language models; and

summing the product for each of the at least two language models.

10. The method according to claim 1, wherein the text data for training the plurality of sets of coefficients is different than language model text data used to train the at least two language models.

11. A method for combining language model scores generated by at least two language models comprised in an Automatic Speech Recognition (ASR) system, said method comprising the steps of:

generating a list of most likely words for a current word in a word sequence uttered by a speaker, and acoustic scores corresponding to the most likely words;

computing language model scores for each of the most likely words in the list, for each of the at least two language models;

respectively and dynamically determining a weight vector to be used to combine the language model scores of each of the most likely words in the list based on a context of the current word, the weight vector comprising n-weights, wherein n equals a number of language models in the system, and each of the n-weights depends upon history n-gram counts; and

respectively combining the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the weight vector determined therefor.

12. The method according to claim 11, wherein said determining step comprises the steps of:

dividing text data for training a plurality of weight vectors into partitions, depending on words counts corresponding to each of the at least two language models; and

for each of the most likely words in the list, dynamically selecting the weight vector from among the plurality of weight vectors so as to maximize the likelihood of the text data with respect to the at least two language models.

13. The method according to claim 11, wherein the at least two language models comprise a first and a second language model, and said dividing step comprises the step of grouping, in a same partition, word triplets $w_1w_2w_3$ which have a count for the word pair w_1w_2 in the first language model greater than the count for the word pair w_1w_2 in the second language model.

14. The method according to claim 12, wherein said selecting step comprises the step of applying the Baum Welch iterative algorithm to the plurality of weight vectors.

15. The method according to claim 11, further comprising the step of, for each of the most likely words in the list, combining an acoustic score and the composite score to identify a group of most likely words to be further processed.

16. The method according to claim 12, wherein the partitions are independent from the at least two language models.

17. The method according to claim 11, wherein each of the plurality of weight vectors comprise a set of coefficients, and said combining step comprises the steps of:

for each of the most likely words in the list,

multiplying a coefficient corresponding to a language model by a language model score corresponding to the language model to obtain a product for each of the at least two language models; and

summing the product for each of the at least two language models.

18. The method according to claim 11, wherein the text data for training the plurality of sets of coefficients is different than language model text data used to train the at least two language models.

19. A combining system for combining language model scores generated by at least two language models comprised in an Automatic Speech Recognition (ASR) system, the ASR system having a fast match for generating a list of most likely words for a current word in a word sequence uttered by a speaker and acoustic scores corresponding to the most likely words, said combining system comprising:

a language model score computation device adapted to compute language model scores for each of the most likely words in the list, for each of the at least two language models;

a selection device adapted to respectively and dynamically select a weight vector to be used to combine the language model scores of each of the most likely words in the list based on a context of the current word, the weight vector comprising n-weights, wherein n

equals a number of language models in the system, and each of the n-weights depends upon history n-gram counts; and

a combination device adapted to respectively combine the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the weight vector selected therefor.

20. The combining system according to claim 19, further comprising a dividing device adapted to divide text data for training a plurality of weight vectors into partitions, depending on words counts corresponding to each of the at least two language models.

21. The combining system according to claim 20, wherein said selection device is further adapted, for each of the most likely words in the list, to dynamically select the weight vector from among the plurality of weight vectors so as to maximize the likelihood of the text data with respect to the at least two language models.

22. The combining system according to claim 19, wherein the at least two language models comprise a first and a second language model, and said dividing device is further adapted to group, in a same partition, word triplets $w_1w_2w_3$ which have a count for the word pair w_1w_2 in the first language model greater than the count of the word pair w_1w_2 in the second language model.

23. The combining system according to claim 20, wherein the partitions are independent from the at least two language models.

24. The combining system according to claim 19, wherein each of the plurality of weight vectors comprise a set of coefficients, and said combining device is adapted, for each of the most likely words in the list, to multiply a coefficient corresponding to a language model by a language model score corresponding to the language model to obtain a product for each of the at least two language models, and to sum the product for each of the at least two language models.

Evidence Appendix

None

Related Procedings Appendix

None

PTO/SB/17 (12-04)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL
For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 500.00

Complete if Known	
Application Number	09/782,434
Filing Date	February 13, 2001
First Named Inventor	Martin Franz
Examiner Name	Jackson, Jakieda R.
Art Unit	2655
Attorney Docket No.	YOR92001-0011US1(8728-481)

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): ____

☒ Deposit Account Deposit Account Number: 50-0510/IBM Deposit Account Name: IBM (Yorktown Heights)

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

☒ Charge fee(s) indicated below
☐ Charge fee(s) indicated below, **except for the filing fee**
☒ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17
☐ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES

Application Type	FILING FEES Fee ($)	FILING FEES Small Entity Fee ($)	SEARCH FEES Fee ($)	SEARCH FEES Small Entity Fee ($)	EXAMINATION FEES Fee ($)	EXAMINATION FEES Small Entity Fee ($)	Fees Paid ($)
Utility	300	150	500	250	200	100	
Design	200	100	100	50	130	65	
Plant	200	100	300	150	160	80	
Reissue	300	150	500	250	600	300	
Provisional	200	100	0	0	0	0	

2. EXCESS CLAIM FEES

Fee Description	Fee ($)	Small Entity Fee ($)
Each claim over 20 or, for Reissues, each claim over 20 and more than in the original patent	50	25
Each independent claim over 3 or, for Reissues, each independent claim more than in the original patent	200	100
Multiple dependent claims	360	180

Total Claims ____ - 20 or HP = **Extra Claims** ____ x **Fee ($)** ____ = **Fee Paid ($)** ____

HP = highest number of total claims paid for, if greater than 20

Multiple Dependent Claims **Fee ($)** ____ **Fee Paid ($)** ____

Indep. Claims ____ - 3 or HP = **Extra Claims** ____ x **Fee ($)** ____ = **Fee Paid ($)** ____

HP = highest number of independent claims paid for, if greater than 3

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets ____ - 100 = **Extra Sheets** 0 / 50 = **Number of each additional 50 or fraction thereof** ____ (round **up** to a whole number) x **Fee ($)** ____ = **Fee Paid ($)** ____

4. OTHER FEE(S) **Fees Paid ($)**

Non-English Specification, $130 fee (no small entity discount) ____

Other: Filing a brief in support of an appeal $500.00

SUBMITTED BY

Signature	Koon Hon Wong (signature)	Registration No. (Attorney/Agent) 48,459	Telephone 516 692-8888
Name (Print/Type)	Koon Hon Wong		Date January 3, 2005

Attorney Docket No.: YOR92001-011US1 (8728-481)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Franz et al. Serial No.: 09/782,434

Filed: February 13, 2001

For: **DYNAMIC LANGUAGE MODEL MIXTURES WITH HISTORY-BASED BUCKETS**

Mail Stop Appeal Brief-Patents
Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

CERTIFICATE OF MAILING

I hereby certify that the following correspondence is being deposited with the United States Postal Service as first class mail, postpaid in an envelope, addressed to the: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on **January 3, 2005**.

1. Appellant's Appeal Brief
2. Fee Transmittal (in triplicate)
3. Return Postcard
4. Certificate of Mailing

Dated: January 3, 2005

Koon Hon Wong

Koon Hon Wong

F. CHAU & ASSOCIATES, LLC
130 Woodbury Road
Woodbury, New York 11797
Tel: (516) 692-8888
Fax: (516) 692-8889

OIPE JAN 0 5 2005 PATENT & TRADEMARK OFFICE

PTO/SB/17 (12-04)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL
For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 500.00

Complete if Known	
Application Number	09/782,434
Filing Date	February 13, 2001
First Named Inventor	Martin Franz
Examiner Name	Jackson, Jakieda R.
Art Unit	2655
Attorney Docket No.	YOR92001-0011US1(8728-481)

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): ____

[X] Deposit Account Deposit Account Number: 50-0510/IBM Deposit Account Name: IBM (Yorktown Heights)

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

[X] Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17 ☐ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES

Application Type	FILING FEES Fee ($)	FILING FEES Small Entity Fee ($)	SEARCH FEES Fee ($)	SEARCH FEES Small Entity Fee ($)	EXAMINATION FEES Fee ($)	EXAMINATION FEES Small Entity Fee ($)	Fees Paid ($)
Utility	300	150	500	250	200	100	
Design	200	100	100	50	130	65	
Plant	200	100	300	150	160	80	
Reissue	300	150	500	250	600	300	
Provisional	200	100	0	0	0	0	

2. EXCESS CLAIM FEES

Fee Description	Fee ($)	Small Entity Fee ($)
Each claim over 20 or, for Reissues, each claim over 20 and more than in the original patent	50	25
Each independent claim over 3 or, for Reissues, each independent claim more than in the original patent	200	100
Multiple dependent claims	360	180

Total Claims ____ - 20 or HP = **Extra Claims** ____ x **Fee ($)** ____ = **Fee Paid ($)** ____

HP = highest number of total claims paid for, if greater than 20

Multiple Dependent Claims **Fee ($)** ____ **Fee Paid ($)** ____

Indep. Claims ____ - 3 or HP = **Extra Claims** ____ x **Fee ($)** ____ = **Fee Paid ($)** ____

HP = highest number of independent claims paid for, if greater than 3

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets ____ - 100 = **Extra Sheets** 0 / 50 = **Number of each additional 50 or fraction thereof** ____ (round up to a whole number) x **Fee ($)** ____ = **Fee Paid ($)** ____

4. OTHER FEE(S) **Fees Paid ($)**

Non-English Specification, $130 fee (no small entity discount) ____

Other: Filing a brief in support of an appeal $500.00

SUBMITTED BY			
Signature	Koon Hon Wong	Registration No. (Attorney/Agent) 48,459	Telephone 516 692-8888
Name (Print/Type)	Koon Hon Wong		Date January 3, 2005

This collection of information is required by 37 CFR 1.136. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.